

Analysis of bibliometric indicators for individual scholars in a large data set

Filippo Radicchi^{1,*} and Claudio Castellano^{2,†}

¹*Departament d'Enginyeria Química, Universitat Rovira i Virgili,
Av. Paisos Catalans 26, 43007 Tarragona, Catalunya, Spain*

²*Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, I-00185 Roma, Italy
Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, I-00185, Roma, Italy*

Citation numbers and other quantities derived from bibliographic databases are becoming standard tools for the assessment of productivity and impact of research activities. Though widely used, still their statistical properties have not been well established so far. This is especially true in the case of bibliometric indicators aimed at the evaluation of individual scholars, because large-scale data sets are typically difficult to be retrieved. Here, we take advantage of a recently introduced large bibliographic data set, Google Scholar Citations, which collects the entire publication record of individual scholars. We analyze the scientific profile of more than 30,000 researchers, and study the relation between the h -index, the number of publications and the number of citations of individual scientists. While the number of publications of a scientist has a rather weak relation with his/her h -index, we find that the h -index of a scientist is strongly correlated with the number of citations that she/he has received so that the number of citations can be effectively be used as a proxy of the h -index. Allowing for the h -index to depend on both the number of citations and the number of publications, we find only a minor improvement.

I. INTRODUCTION

Bibliographic databases play nowadays a crucial role in modern science. Citation numbers, or other measures derived from bibliographic data, are commonly used as quantitative indicators for the impact of research activities. Citation analysis has been criticized [1, 29, 30], and the true meaning of a citation can be very different from context to context [5, 17]. Despite these objections, the use of citations is widespread and citation numbers are currently and frequently used for assessing the impact of individual scholars [12, 20], journals [15], departments [10], universities and institutions [25]. Especially at the level of individual scientists, numerical indicators based on citation counts are evaluation tools of fundamental importance for decisions about hiring [4] and/or grant awards [6].

Though widely used, numerical indicators based on citation numbers are generally poorly understood [28]. Even in the basic case of citation distributions of papers, where data are easily collectable and analyzable, there is no clear general picture. Depending on the study performed and the data set analyzed, citation distributions have been judged compatible with several possible statistical distributions: power-law functions [11, 41], log-normal distributions [39, 47, 48], stretched exponentials [27, 49], and others. At the same time, however, while researchers have not yet reached an agreement on the precise law governing citation distributions, interesting properties in citation data are nevertheless visible and detectable [38]. Up to now, large-scale statistical analyses have been lim-

ited to the study of citations accumulated by papers [38, 49] or journals [43, 50]. In these cases data are easily collected from the main bibliographic databases available on the market and do not require special filtering procedures. Conversely, the collection of bibliographic data about individual scholars is much more difficult. Simple searches on bibliographic databases are generally unable to produce clean data sets because of evident problems related to the proper disambiguation of scientists. All studies conducted so far have been therefore limited either to small sets of scientists [3, 6, 9, 20, 32–36, 42, 44], or to data sets subjected to disambiguation problems [7, 28, 40].

Here, we take advantage of a data set composed of more than 30,000 [about two orders of magnitude larger than those used by Bar-Ilan [3], Bornmann et al [6], Costas and Bordons [9], Hirsch [20], Petersen et al [32, 33, 34, 35], Redner [42], Schreiber et al [44]] individual scientific profiles. The data set is rather clean because profiles are directly managed by scientists themselves, who are interested in providing correct information about the outcome of their research activity. We perform an initial exploratory analysis of this data set, and show that the main basic quantities used in research evaluation exercises obey well-defined statistical distributions. We then use the data set to investigate (on a scale more than 10 times larger than previous studies) the relation between the h -index and other simple bibliometric indicators.

II. DATA SET

We collected the scientific profiles of 89,786 scientists from Google Scholar Citations (GSC, scholar.google.com/citations) database. The profile of each scholar reports the entire publication record of the scientist, including the year of publication

*Electronic address: f.radicchi@gmail.com

†Electronic address: claudio.castellano@roma1.infn.it

and the number of citations accumulated by each publication according to the Google Scholar database [for studies about differences in the quantification of bibliometric indicators of individual scientists between Google Scholar and other popular bibliographic databases see [2, 3, 22, 23, 31]]. The profile is owned and managed by scientists themselves, who can delete and add publications, even merge two publications if considered initially as different in the database, and thus provides a clean source of information. Scientists are requested to validate their profile by providing their academic email address. This validation ensures that the profile is actually managed by the scientist. Finally, each scientist is required to provide a set of keywords which identify the research fields in which the scientist is active.

Data have been collected between June 29 and July 4, 2012. Number of publications and citations reflect therefore the research activity performed until that time. We used an iterative procedure consisting in downloading the entire set of authors using an initial keyword (we used “network science”), adding the other keywords used by these scientists, downloading the profile of new scientists that are using these new keywords, and so on, until we were able to discover neither new scientists nor new keywords. In total, we were able to identify 67,648 different keywords (see Fig. 1 for a word cloud of the most common keywords). It is important to notice that the database is in rapid evolution and growth. For example, we used the same procedure described above to download data in March 2012, and at that time the data set was composed of 49,365 scientists and 38,679 keywords.

In order to be sure about the information provided by users, have a better control of the publication record of individual scientists, and include only scholars with a sufficiently long period of activity and sufficiently large number of publications, we filter the data set with the following restrictions:

1. We restrict our analysis to the 83,897 scientists who validated their profile with an academic email.
2. We delete from the data set publications that were published before year 1945. This was necessary in order to exclude from the data set papers whose year of publication is wrong in GCS. Note also that scientists with first publication before year 1945, if still active and with a validated profile, would have academic ages longer than 67 years.
3. We further restrict the attention to scientists with at least 20 publications and career length longer than or equal to 5 years (the academic age or career length is measured as the difference between the publication years of the first paper of a scholar and year 2012).

In the rest of the paper, we will present the result of

the analysis based on a total of 35,136 scholars whose research profile satisfies the aforementioned conditions.

III. RESULTS

A. General properties of the data set

We first investigate general properties of the population in our data set. In Fig. 2A, we show the composition of the population in terms of academic age. The probability density function (pdf) $P(A)$ of the career length A can be reasonably well described (by graphical inspection, although the measured p -value does not support a good statistical compatibility) by a log-normal distribution

$$P(A) = \frac{1}{A \sqrt{\pi \sigma^2}} e^{-[\log(A) - \mu]^2 / (2\sigma^2)} , \quad (1)$$

and the best estimate of parameters (obtained with least square fit) of the distribution are $\hat{\mu}_A = 2.89$ and $\hat{\sigma}_A = 0.51$ (the suffix A is used to indicate that the parameters have been calculated for the academic age A).

The number of citations C received by each scientist (C is the sum of all citations accumulated by all papers written by an author) is well fitted by (see Fig. 2B)

$$P(C) = \frac{1}{C \tau} e^{-z - e^{-z}} , \quad (2)$$

where $z = \frac{\log(C) - \nu}{\tau}$, and the best estimates of the parameters are $\hat{\nu}_C = 6.42$ and $\hat{\tau}_C = 1.22$. Eq. 2 is a generalization to the logarithms of the well-known Gumbel function that usually appears in the description of the statistics of extreme values.

In Figs. 2C and D, we report the pdfs of the number of publications N and the h -index, respectively. In these cases, we tried to fit the distributions with both Eqs. 1 and 2, but none of them was able to describe entirely the pdfs obtained with data. It is, however, interesting to note that the pdf of the number of publications per author is neither a strict decreasing function nor a power-law function as often assumed in the literature [13], but instead the pdf in Fig. 2C shows a clear peak and a decay faster than a power-law at large values of N .

In general, the results presented in Fig. 2 depend on the choices we made in the selection of the authors. For example, the peak position of the $P(A)$ [Fig. 2A] moves from $A = 14$ to $A = 6$ if we remove the restriction on the minimal number of publications needed to enter in the sample. Similar considerations are also valid for the other pdfs. On the other hand, our choices do not affect the tail of the pdfs, and more generally their shapes. For example, even if the peak moves to lower values when we include all authors in the data set, the pdf of Fig. 2A still can be reasonably well described by a log-normal distribution.

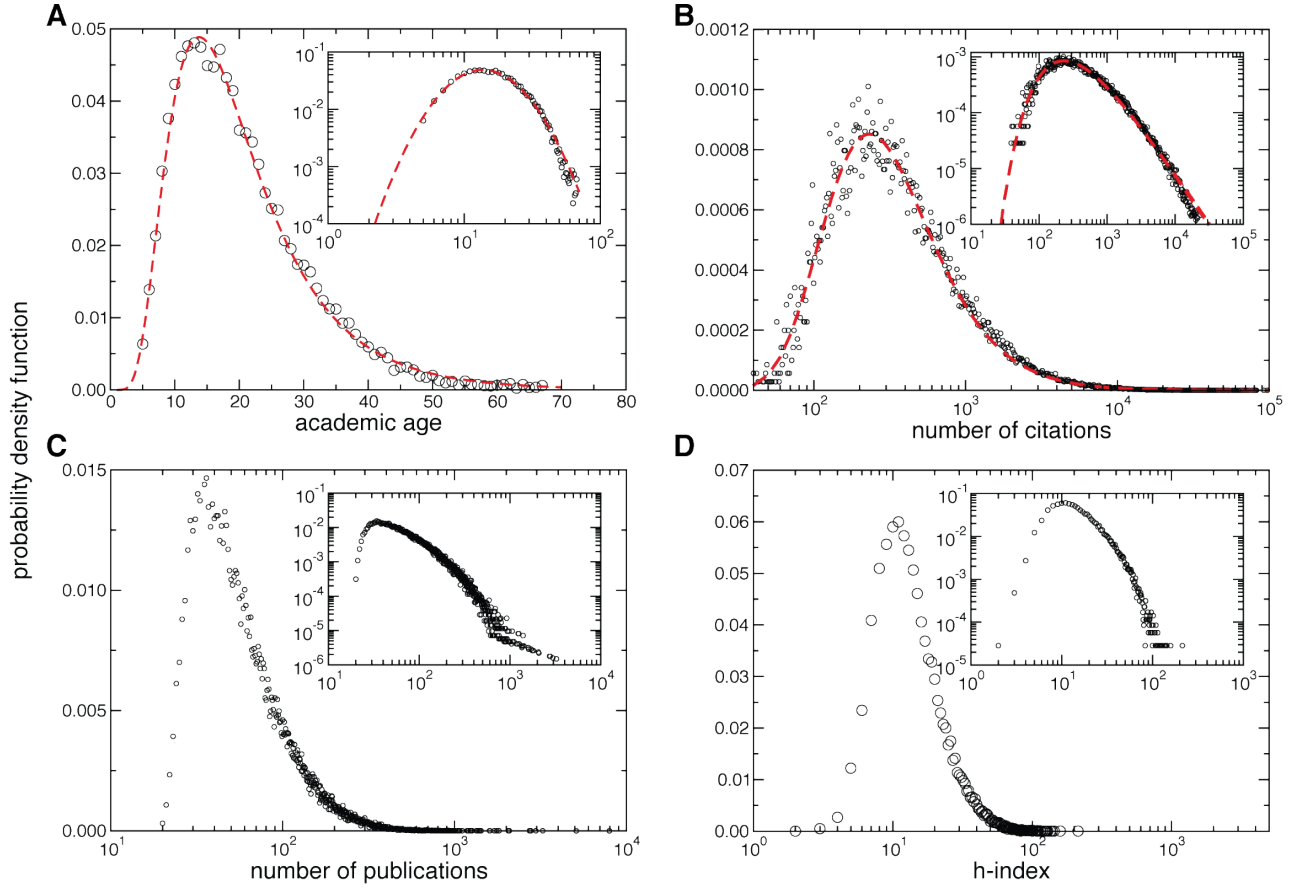


Figure 2: **A.** Probability density function of the academic age of the scientists in the data set. Data are fitted with a log-normal distribution [Eq. 1] with parameters values $\hat{\mu}_A = 2.89$ and $\hat{\sigma}_A = 0.51$ (red dashed line). **B.** Probability density function of the total number of citations received by the scientists in the data set. Data are fitted with a log-Gumbel distribution [Eq. 2] with parameters values $\hat{\nu}_C = 6.42$ and $\hat{\tau}_C = 1.22$ (red dashed line). **C.** Probability density function of the total number of publications produced by the scientists in the data set. **D.** Probability density function of the h -index of the scientists in the data set.

the coefficient of variation, is $\hat{\alpha}_{h,N} = 2.0$. Notice that this value indicates that h depends differently on C and on N , thus contradicting the hypothesis that $f(c)$ is, for all scholars, a pure power-law over its whole range. In Fig. 4B we plot the pdf of the quantity $N^{0.50}/h$, which is approximately fitted by a log-normal distribution with parameters values $\hat{\mu}_{h,N} = -0.64$ and $\hat{\sigma}_{h,N} = 0.39$. Also in this case we find quite a narrow distribution, but the value of the coefficient of variation indicates a worse agreement with data with respect to one found for h vs. C . Indeed, the inset of Fig. 3A shows that the minimum coefficient of variation (corresponding to $\hat{\alpha}_{h,C} = 2.39$) is around 0.19, while the minimum in the inset of Fig. 4A is around 0.45.

Under the power-law assumption for $f(c)$, it is also possible to express h as a function of both N and the average number of citations per paper $\chi = C/N$, obtain-

ing [21, 45]

$$h \sim \chi^{(\alpha_{h,C,N}-1)/\alpha_{h,C,N}} N^{1/\alpha_{h,C,N}} \quad (5)$$

$$\sim C^{(\alpha_{h,C,N}-1)/\alpha_{h,C,N}} N^{(2-\alpha_{h,C,N})/\alpha_{h,C,N}}. \quad (6)$$

Minimizing the coefficient of variation for the quantity $C^{(\alpha_{h,C,N}-1)/\alpha_{h,C,N}} N^{(2-\alpha_{h,C,N})/\alpha_{h,C,N}}/h$, we determine the best estimate of $\alpha_{h,C,N}$ as $\hat{\alpha}_{h,C,N} = 1.70$ (inset of Fig. 5A), implying $h \sim C^{0.41} N^{0.18}$. In the main panel of Fig. 5A we plot, for each author, h vs. $C^{0.41} N^{0.18}$ finding a good agreement with the expected linear behavior. Also in this case we find that the quantities are correlated ($R_{h,C,N} = 0.94$). The pdf of the rescaled index $C^{0.41} N^{0.18}/h$ is well peaked (Fig. 5B), and the distribution can be reasonably well fitted by a log-Gumbel distribution with best-fit parameters $\hat{\nu}_{h,C,N} = 0.79$ and $\hat{\tau}_{h,C,N} = 0.14$.

As the relation $h \sim C^{0.41} N^{0.18}$ shows, the dependence of h on C is much stronger than the one on N . The comparison between the minimum coefficients of variation measured for the three scaling assumptions (Eqs. 3,

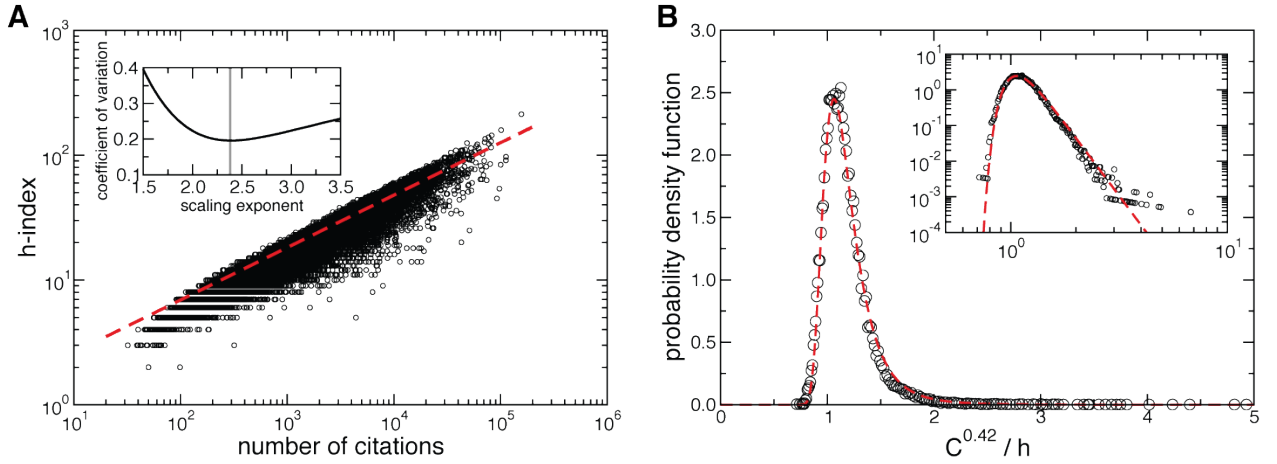


Figure 3: **A.** Relation between the h -index and the number of citations C for each scientist in our data set. We fit data with a power-law function [Eq. 3], whose best estimate of the exponent equals $1/\hat{\alpha}_{h,C} = 1/2.39 = 0.42$ (dashed line). **B.** Probability density function of the quantity $C^{0.42}/h$. This function is fitted by a log-Gumbel distribution [Eq. 2] with parameter values $\hat{\nu}_{h,C} = 0.08$ and $\hat{\tau}_{h,C} = 0.14$ (dashed line). The inset shows the same as the main plot but in a double-logarithmic scale.

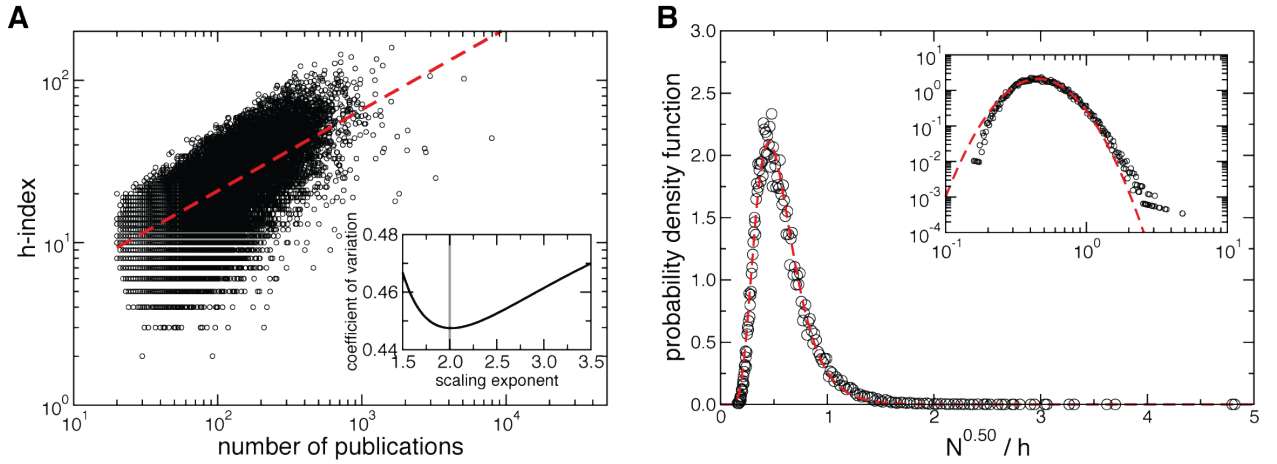


Figure 4: **A.** Relation between the h -index and the number of publications N for each scientist in our data set. We fit data with a power-law function, whose best estimate of the exponent equals $\hat{\alpha}_{h,N} = 0.50$ (dashed line). **B.** Probability density function of the quantity $N^{0.50}/h$. This function is fitted by a log-normal distribution with parameter values $\hat{\mu}_{h,N} = -0.64$ and $\hat{\sigma}_{h,N} = 0.39$ (dashed line). The inset shows the same as the main plot but in a linear-logarithmic scale.

4 and 6) indicates that allowing for a dependence on both C and N leads only to a marginal improvement over considering only the dependence on C (the coefficient of variation changes only from 0.19 to 0.18) while a dependence only on N performs definitely worse. The presence of a term dependent on N in Eq. 6 brings only a little improvement and leaves the exponent of the dependence on C practically unaltered (0.41 *vs.* 0.42).

IV. CONCLUSIONS

Statistical analysis of bibliometric indicators devoted to the evaluation of individual scholars is usually diffi-

cult because of the lack of large and clean data sets describing accurately the publication records of researchers. This is a general problem that regards every bibliographic database available on the market, and is also the main reason for which the studies performed so far on the characterization of the bibliographic profile of individual scientists have been rarely based on more than 1,000 individuals. In recent years, however, some main bibliographic databases have started to allow individual scientists to freely manage their publication profiles with specially designed on-line tools. This is the case of the recently created ResearcherID by Thomson Reuters (<http://www.researcherid.com>), Mendeley profiles (<http://www.mendeley.com>) and also

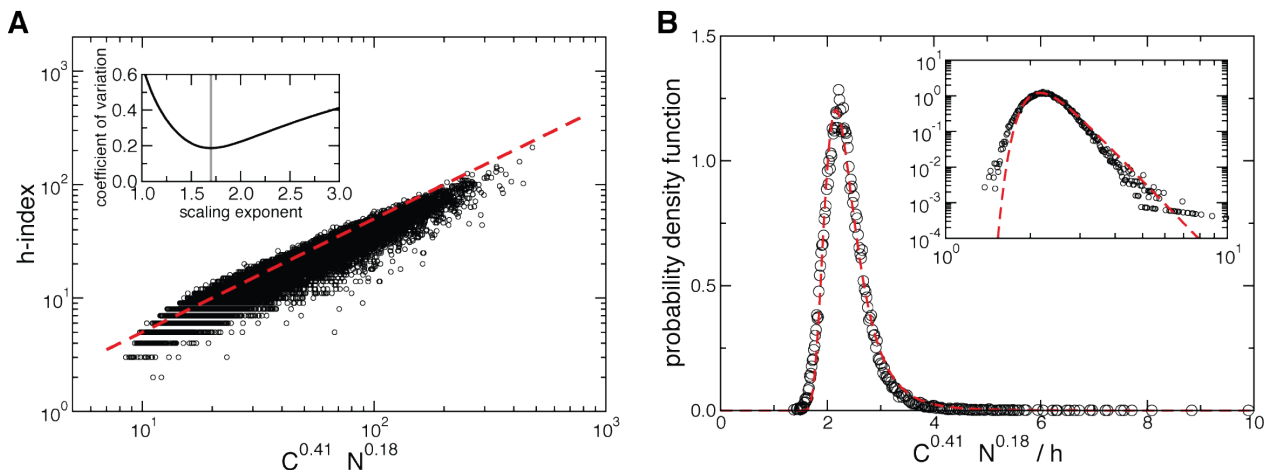


Figure 5: **A.** Relation between the h -index, the number of citations C and the number of publications N for each scientist in our data set [Eq. 6]. The best estimate of the exponent equals $\hat{\alpha}_{h,C,N} = 1.70$ (see inset). Effectively, when h is plotted against $C^{0.41} N^{0.18}$ (i.e., Eq. 6 for $\alpha_{h,C,N} = \hat{\alpha}_{h,C,N}$), we recover a good linear behavior (dashed line). Probability density function of the quantity $C^{0.41} N^{0.18}/h$. This function is fitted by a log-Gumbel distribution with parameter values $\hat{\nu}_{h,C,N} = 0.79$ and $\hat{\tau}_{h,C,N} = 0.14$ (dashed line). The inset shows the same as the main plot but in a double-logarithmic scale.

of Google Scholar Citations (scholar.google.com/citations). In all these online administration systems, scholars manage directly their profiles by adding, deleting and correcting their publication records, and thus the information provided can be considered accurate because it is in the interest of researchers to provide an accurate and up-to-date source of information regarding their research production. Here, we took advantage of the entire data set of Google Scholar Citations as of June 29, 2012. The data set is composed of more than 30,000 individual scholars working in research institutions worldwide. Although Google Scholar Citations represents a relatively clean set of data describing the academic records of individual researchers, it is important to stress that our set of data does not represent a random sample of researchers because the presence of a researcher in the system is subjected to various types of factors that, as a matter of fact, bias the sample. First, our data set is mainly composed of relatively young scientists (Fig. 2A) who are able to create a profile, validate it with their email, and manage it. Second, the profiles that compose the data set are certainly those of scholars who want to promote their research with the use of modern information-technology tools. Finally, although the scientists present in our data set have various fields of expertise, some scientific disciplines are clearly over-represented and others under-represented (see Fig. 1). We would like to further emphasize that the entire data set analyzed here is clearly subjected to all the limitations of the Google Scholar database (eventual presence of fake publications, duplication of citations, etc.) that have been deeply studied in the literature [18, 24, 26].

Taking into account the formerly mentioned limitations of our data set, here we provided an exploratory analy-

sis on some basic statistics for single authors and focus in some detail on the relation between the h -index, the number of publications and the number of citations of individual scientists. We found three main results:

1. The h -index h is strongly correlated with the total number of citations C received by a scientist with his/her own scientific production. In particular, we find $h \sim C^{0.42}$ in qualitative agreement with the early hypothesis by Hirsch [20] validated empirically on small data sets [37, 42, 46].
2. The h -index is also shown to be correlated with the number of publications N , but this relation is much less precise than the one observed for C .
3. It is possible to combine both dependencies into a single power-law relation $h \sim C^{0.41} N^{0.18}$. This law, however, provides only a slight improvement with respect to the power-law relation that connects only h and C .

Our results represent a large-scale validation of formerly postulated conjectures. While the exact values of the measured power-law exponents might be data set dependent, we believe that the main message has a validity that goes beyond the data analyzed here: the total number of citations C received by a scientist can be used as an effective proxy of his/her h -index.

The fact that h is strongly correlated with C and much more weakly with the total number of publications N is evidence that the distribution $f(c)$ of citations accrued by publications of a single researcher is not a pure power-law over its whole range. Please note that we do not exclude the possibility that this fact is a consequence of the sample used in our analysis, where scholars have academic age typically shorter than the one of an average

researcher and thus individual citation distributions may have not yet reached as sufficient level of stationarity. Our results, however, call for more work to better characterize and understand the activity and citations profile of individual scholars, their common features and their

variations. The large data set provided by Google Scholar Citations constitutes an ideal tool for this endeavor. The present study represents only a first attempt to scratch the surface of such a treasure trove.

-
- [1] Adler R, Ewing J, Taylor P (2009) Citation Statistics. *Statistical Science* 24(1):1–14
 - [2] Alonso S, Cabrerizo F, Herrera-Viedma E, F H (2009) h-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics* 3(4):273–289
 - [3] Bar-Ilan J (2008) Which h-index?—a comparison of WOS, Scopus and Google Scholar. *Scientometrics* 74(2):257–271
 - [4] Bornmann L, Daniel HD (2006) Selecting scientific excellence through committee peer review - A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics* 68(3):427–440
 - [5] Bornmann L, Daniel HD (2008) What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation* 64(1):45–80
 - [6] Bornmann L, Wallon G, Ledin A (2008) Does the Committee Peer Review Select the Best Applicants for Funding? An Investigation of the Selection Process for Two European Molecular Biology Organization Programmes. *PLoS ONE* 3(10):e3480
 - [7] Cabanac G (2013) Experimenting with the partnership ability φ -index on a million computer scientists. *Scientometrics*. To appear.
 - [8] Costas R, Bordons M (2007) The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics* 1(3):193–203
 - [9] Costas R, Bordons M (2008) Is g-index better than h-index? an exploratory study at the individual level. *Scientometrics* 77(2):267–288
 - [10] Davis P, Papanek GF (1984) Faculty Ratings of Major Economics Departments by Citations. *The American Economic Review* 74(1):225–230
 - [11] De Solla Price DJ (1965) Networks of Scientific Papers. *Science* 149(3683):510–515
 - [12] Egghe L (2006) Theory and practise of the g-index. *Scientometrics* 69(1):131–152
 - [13] Egghe L (2010) The hirsch index and related impact measures. *Annual Review of Information Science and Technology* 44(1):65–114
 - [14] Egghe L, Rousseau R (2006) An informetric model for the hirsch-index. *Scientometrics* 69(1):121–129
 - [15] Garfield E (1998) The impact factor and using it correctly. *Der Unfallchirurg* 101(6):413–414
 - [16] Glänzel W (2006) On the h-index - a mathematical approach to a new measure of publication activity and citation impact. *Scientometrics* 67(2):315–321
 - [17] Hartley J (2012) To cite or not to cite: author self-citations and the impact factor. *Scientometrics* 92(2):313–317
 - [18] Harzing AWK, van der Wal R (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics* 8(1):61–73
 - [19] Hendricks WA, Robey KW (1936) The Sampling Distribution of the Coefficient of Variation. *The Annals of Mathematical Statistics* 7(3):129–132
 - [20] Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102(46):16,569–16,572
 - [21] Iglesias J, Pecharromán C (2007) Scaling the h-index for different scientific ISI fields. *Scientometrics* 73(3):303–320
 - [22] Jacsó P (2005) As we may search – comparison of major features of web of science, scopus and Google Scholar citation-based and citation-enhanced databases. *Current Science* 89(9):1537–1547
 - [23] Jacsó P (2005) Visualizing overlap and rank differences among web-wide search engines. *Online Information Review* 29(5):554–560
 - [24] Jacsó P (2010) Metadata mega mess in Google Scholar. *Online Information Review*, 34(1):175–191.
 - [25] Kinney AL (2007) National scientific facilities and their science impact on nonbiomedical research. *Proceedings of the National Academy of Sciences of the United States of America* 104(46):17,943–17,947
 - [26] Labbé C (2011) Ike Antkare, one of the great stars in the scientific firmament. *ISSI newsletter*, 6(2):48–52.
 - [27] Laherrère J, Sornette D (1998) Stretched exponential distributions in Nature and Economy: “Fat tails” with characteristic scales. *European Physical Journal B* 2(4):525–539
 - [28] Lehmann S, Jackson AD, Lautrup BE (2006) Measures for measures. *Nature* 444(7122):1003–1004
 - [29] MacRoberts MH, MacRoberts BR (1989) Problems of citation analysis: A critical review. *Journal of the American Society for Information Science* 40(5):342–349
 - [30] MacRoberts MH, MacRoberts BR (1996) Problems of citation analysis. *Scientometrics* 36(3):435–444
 - [31] Meho LI, Yang K (2007) Impact of data sources on citation counts and rankings of lis faculty: Web of science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology* 58(13):2105–2125
 - [32] Petersen AM, Jung Ws, Yang Js, Stanley HE (2010) Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proceedings of the National Academy of Sciences* 108(1):18–23
 - [33] Petersen AM, Wang F, Stanley HE (2010) Methods for measuring the citations and productivity of scientists across time and discipline. *Physical Review E* 81(3):1–9
 - [34] Petersen AM, Stanley HE, Succi S (2011) Statistical regularities in the rank-citation profile of scientists. *Scientific reports* 1:181
 - [35] Petersen AM, Riccaboni M, Stanley HE, Pammolli F

- (2012) Persistence and uncertainty in the academic career. *Proceedings of the National Academy of Sciences* 109(14):5213–5218
- [36] Pratelli L, Baccini A, Barabesi L, Marcheselli M (2012) Statistical Analysis of the Hirsch Index. *Scandinavian Journal of Statistics*, 39(4):681–694
- [37] van Raan AFJ (2006) Comparison of the hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics* 67(3):491–502
- [38] Radicchi F, Castellano C (2012) A Reverse Engineering Approach to the Suppression of Citation Biases Reveals Universal Properties of Citation Distributions. *PLoS ONE* 7(3):e33,833
- [39] Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America* 105(45):17,268–17,272
- [40] Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. *Physical Review E* 80(5):056,103
- [41] Redner S (1998) How popular is your paper? An empirical study of citation distribution. *European Physical Journal B* 4(2):131–134
- [42] Redner S (2010) On the meaning of the h-index. *Journal of Statistical Mechanics: Theory and Experiment* (3):L03,005
- [43] Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America* 105(4):1118–1123
- [44] Schreiber M, Malesios C, S P (2011) Categorizing h-index variants. *Research Evaluation* 21(3):397–409
- [45] Schubert A, Glänzel W (2007) A systematic analysis of hirsch-type indices for journals. *Journal of Informetrics* 1(3):179 – 184
- [46] Spruit HC (2012) The relative significance of the H-index. *ArXiv e-prints* 1201.5476
- [47] Stringer MJ, Sales-Pardo M, Amaral LAN (2008) Effectiveness of Journal Ranking Schemes as a Tool for Locating Information. *PLoS ONE* 3(2):e1683
- [48] Stringer MJ, Sales-Pardo M, Amaral LAN (2010) Statistical Validation of a Global Model for the Distribution Published in a Scientific Journal. *Journal of the American Society for Information Science* 61(7):1377–1385
- [49] Wallace ML, Larivière V, Gingras Y (2008) Modeling a Century of Citation Distributions. *Journal of Informetrics* 3(4):296–303
- [50] West J, Bergstrom T, Bergstrom CT, Road HP, Fe S (2010) Big Macs and Eigenfactor Scores : Don’t Let Correlation Coefficients Fool You. *Journal of the American Society for Information Science* 61(2008):1800–1807